An Interop Publication

ConneXions®

The Interoperability Report

November 1991 Volume 5, No. 11

ConneXions — The Interoperability Report tracks current and emerging standards and technologies within the computer and communications industry.

In this issue:

ConneXions is published monthly by Interop, Inc., 480 San Antonio Road, Suite 100, Mountain View, CA 94040, USA. 415-941-3399. Fax: 415-949-1779. Toll-free: 1-800-INTEROP.

ISSN 0894–5926

From the Editor

After two special issues focusing on INTEROP 91 Fall, it is time to return to "normal" and catch up with some of the topics *not* directly related to the show which got pushed aside in the last couple of months. Of course, we will return to INTEROP 91 Fall in a future issue (most likely December 1991) with reports and pictures, so stay tuned.

Our July issue, subtitled "The Changing Face of the Internet," looked at a number of new and interesting applications of Internet technology. This month, we continue this thread with a look at WAIS, Multimedia Mail, and Resource Discovery.

The *Wide Area Information Server* (WAIS, pronounced "ways") project is an experimental venture seeking to determine whether current technologies can be used to make profitable end-user full-text information systems. Our first article, written by Brewster Kahle and Art Medlar, discusses the design and implementation of the prototype WAIS system.

Multimedia mail systems have actually been in use on the Internet and elsewhere for many years. However, no multimedia mail technology has reached critical mass, due in part to the variety of interchange standards and systems in use. Nathaniel Borenstein of Bellcore gives a brief summary of the state of the art in multimedia mail systems. The article describes a new "bottom-up" approach to multimedia mail, and outlines a vision of a new and better "lowest common denominator" for electronic mail.

In a recent study, researchers at the University of Colorado, involved with the Resource Discovery project, attempted to measure the nature of connectivity to the Internet by sending certain simple "probes" to a statistical sample of host. The reaction to this experiment is the subject of an article by Carl Malamud on page 18. It should be noted that the IAB recently issued a statement—in the form of RFC 1262—on the subject of Internet Measurement. The summary is included below:

"Measurement of the Internet is critical for future development, evolution and deployment planning. Internet-wide activities have the potential to interfere with normal operation and must be planned with care and made widely known beforehand. This document offers guidance to researchers planning Internet measurements. This RFC represents IAB guidance for researchers considering measurement experiments on the Internet. This RFC does not represent a standard for the Internet but the Internet Activities Board strongly urges that Internet users follow the guidelines out of courtesy and professional consideration for the Internet community."

An Information System for Corporate Users: Wide Area Information Servers

by
Brewster Kahle, Thinking Machines Corporation
and
Art Medlar, Scolex Information Systems

Background

To explore text-based information systems for corporate executives, four companies have jointly developed a prototype which gives flexible access to full-text documents. The four participating companies are Dow Jones & Co., with its premier business information sources; Thinking Machines Corporation, with its high-end information retrieval engines; Apple Computer, with its user interface expertise; and KPMG Peat Marwick, with its information-hungry user base.

One of the primary objectives of the project is to allow a user to retrieve personal, corporate, and wide area information through one easy-to-use interface. For example, instead of using Lotus *Magellean*™ for personal information, *Verity Topic*™ for corporate data, and *Dialog*™ for published text, one application can access all three categories of information. The user isn't required to become familiar with several entirely different systems. In addition, since the interface consolidates data from many different sources, they can be manipulated effortlessly, virtually without regard to their origins.

The *Wide Area Information Server* (WAIS, pronounced "ways") project is an experimental venture seeking to determine whether current technologies can be used to make profitable end-user full-text information systems. Fifteen users have been actively using the system for over three months. They have integrated it into their workday routine in much the same way as they have previously integrated spreadsheets and word processors. This preliminary success has convinced us that a WAIS-like system can be a valuable tool for corporate information retrieval. This article discusses the design and implementation of the prototype system.

Introduction

Electronic publishing is the distribution of textual information over electronic networks. It has been emerging as a viable alternative to traditional print publishing as the necessary underlying technologies develop. Among the more essential of these are:

- High Resolution Display Screens
- Reliable, High-Speed Data Communications
- Desktop Publishing Systems
- Inexpensive Data Storage Media

While these technologies have been developed for uses other than electronic publishing, they are the necessary precursors for full-text retrieval systems.

From the user's point of view, there are several problems to be overcome. First, there must be some way of finding and selecting databases from a potentially unlimited pool. Second, although these databases may be organized in different ways, the user should not need to become familiar with the internal configuration of each one. Finally, there must be some practical way of organizing responses on the users machine in order to maintain control over what may become a vast accumulation of data.

In addition, developers are faced with a number of architectural issues. The system must be scalable; that is, it must allow for the future growth of both the complexity and number of clients and servers. It must be secure; each server's data must be protected from corruption, and the privacy of the users must be ensured. Lastly, since an unreliable source is useless in a corporate environment, access must be thoroughly robust.

System overview

The prototype WAIS system takes advantage of current state-of-the-art technology, and presents solutions to all of the above problems. The system is composed of three separate parts: *Clients, Servers,* and the *Protocol* which connects them.

The Client is the user interface, the server does the indexing and retrieval of documents, and the protocol is used to transmit the queries and responses. The client and server are isolated from each other through the protocol. Any client which is capable of translating a users request into the standard protocol can be used in the system. Likewise, any server capable of answering a request encoded in the protocol can be used. In order to promote the development of both clients and servers, the protocol specification is public, as is its initial implementation.

On the client side, questions are formulated as English language questions. The client application then translates the query into the WAIS protocol, and transmits it over a network to a server. The server receives the transmission, translates the received packet into its own query language, and searches for documents satisfying the query. The list of relevant documents are then encoded in the protocol, and transmitted back to the client. The client decodes the response, and displays the results. The documents can then be retrieved from the server.

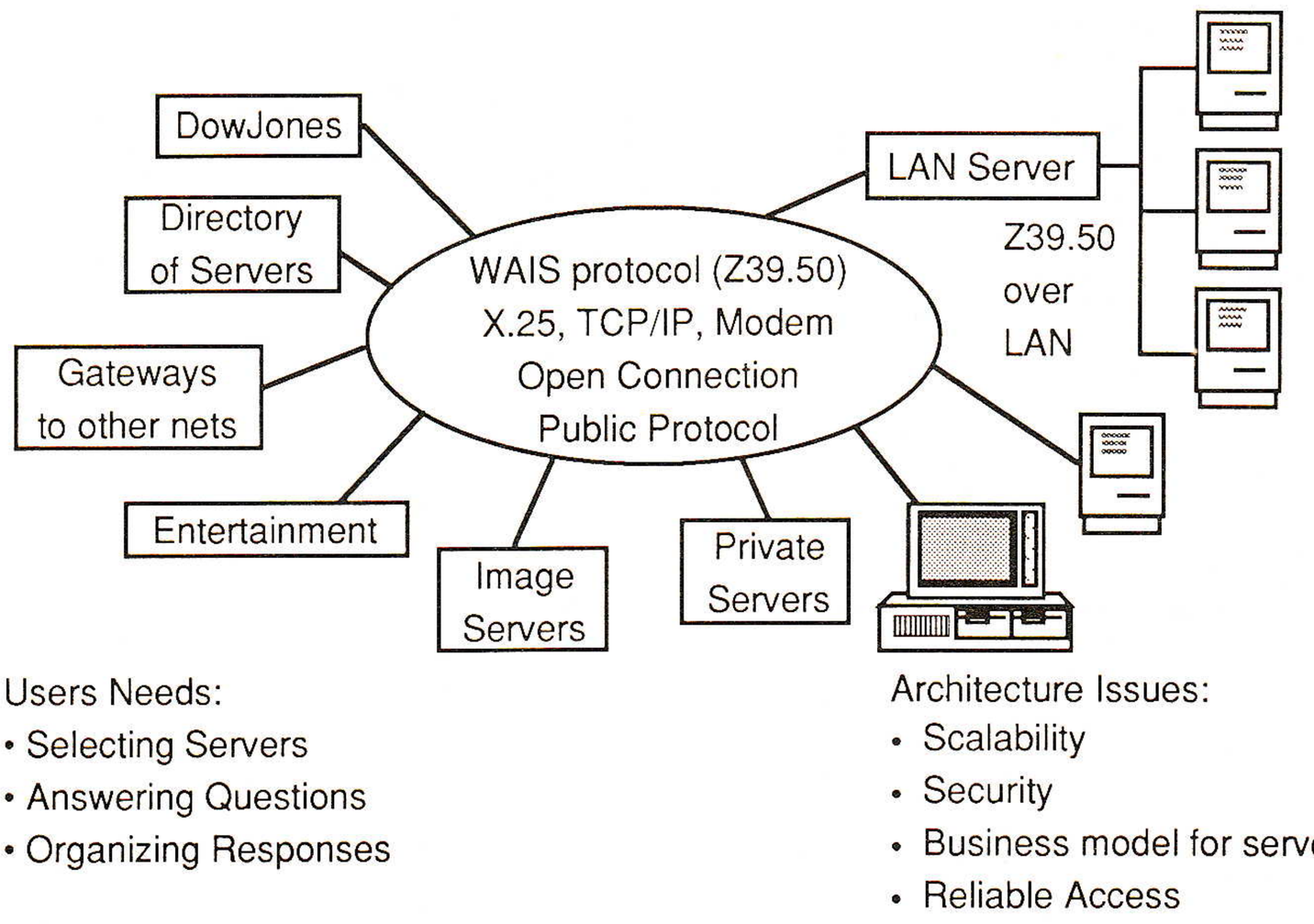


Figure 1: WAIS System Components

Digital researcher

The traditional information research scenario is familiar to anyone who has ever visited a reference desk at a public or corporate library. The client approaches a librarian with a description of needed information. The librarian might ask a few background questions, and then draws from appropriate sources to provide an initial selection of articles, reports, and references.

continued on next page

Wide Area Information Servers *(continued)*

The client then sorts through this selection to find the most pertinent documents. With feedback from these trials, the researcher can refine the materials and even continue to supply the user with a flow of information as it becomes available. Monitoring which articles were useful can help keep the researcher on-track.

The WAIS system is an attempt at automating this interaction: the user states a question in English, and a set of document descriptions come back from selected sources. The user can examine any of the items, be they text, picture, video, sound, or whatever. If the initial response is incomplete or somehow insufficient, the user can refine the question by stating it differently.

In addition, the user may also mark some of the retrieved documents as being "relevant" to the question at hand, and then re-run the search. The server recognizes the marked documents, and attempts to find others which are similar to them. In the present WAIS system, "similar" documents are simply ones which share a large number of common words; however, there is potentially no upper limit on the intelligence of a server in determining what similarity entails. This method of information retrieval is called "relevance feedback." The idea has been around for many years [1] and the first commercial system utilizing it, *DowQuest* [2], was voted Database of the Year by *ONLINE Magazine* in January 1989.

User interfaces

Users interact with the WAIS system through the *Question* interface. The interface may appear different on various implementations: for example, a character display terminal will have a different look than one which is capable of displaying bit-mapped graphics. The key, however, is that the user need only become familiar with one interface which provides access to all available information sources.

The WAIS system, in this first incarnation, was designed to be used by accountants and corporate executives who are relatively untrained in search techniques. Consequently, to aid those users who have neither the time nor desire to learn a special purpose query language, the system uses English language queries augmented with relevance feedback. While the system's servers currently do not extract semantic information from the English queries, they do their best to find and rank articles containing the requested words and phrases. Used in conjunction with relevance feedback, this method of searching has proven to be more than adequate for the types of searches and databases typically encountered.

Several user interfaces are in use or under development at Thinking Machines, Apple Computer, Dow Jones, and elsewhere. As shown on the facing page, a typical search scenario has the following steps:

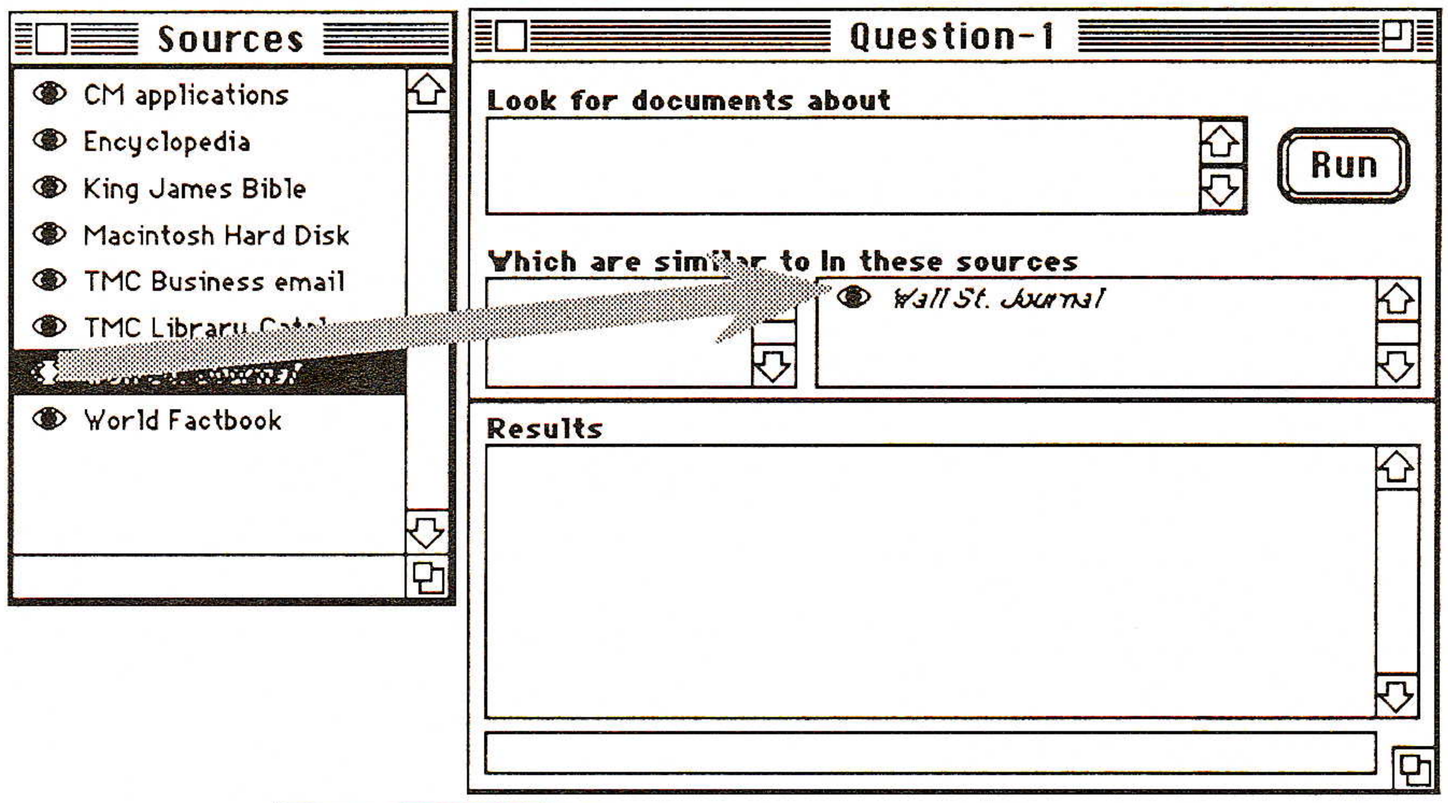
Step 1: Sources are dragged with the mouse into a Question Window. A question can contain multiple sources. When the question is run, it asks for information from each included source.

Step 2: When a query is run, headlines of documents satisfying the query are displayed.

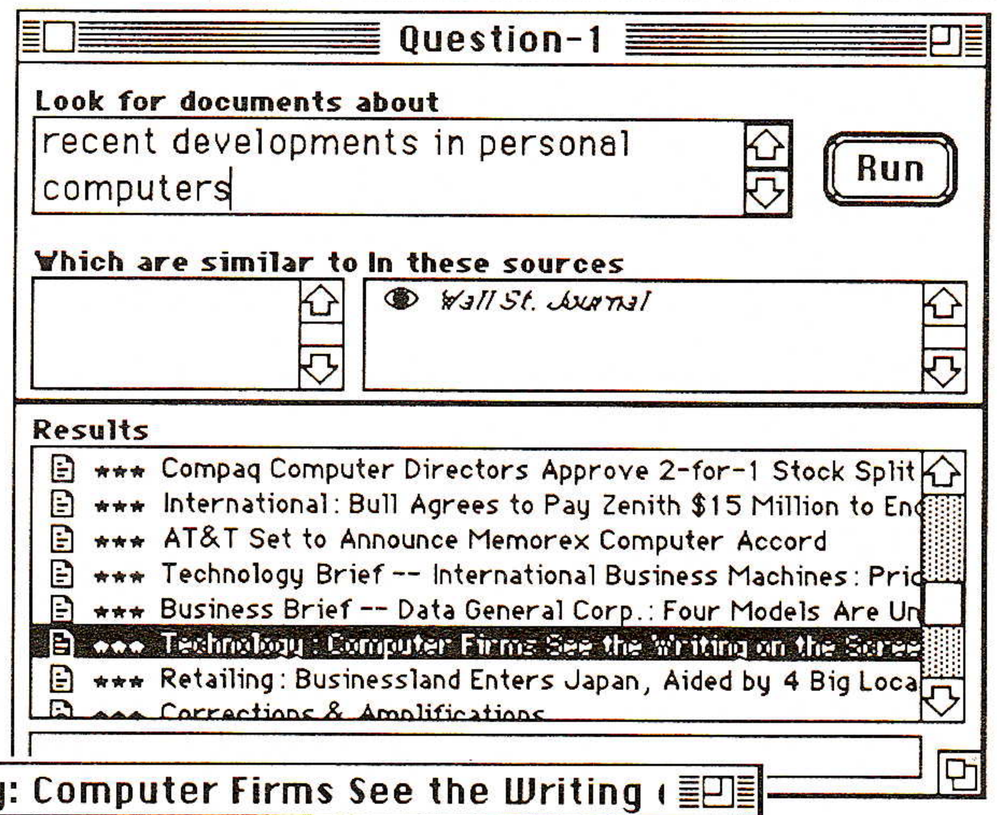
Step 3: With the mouse, the user clicks on any result document to retrieve it.

Step 4: To refine the search, any one or more of the result documents can moved to the "Which are similar to:" box. When the search is run again, the results will be updated to include documents which are "similar" to the ones selected.

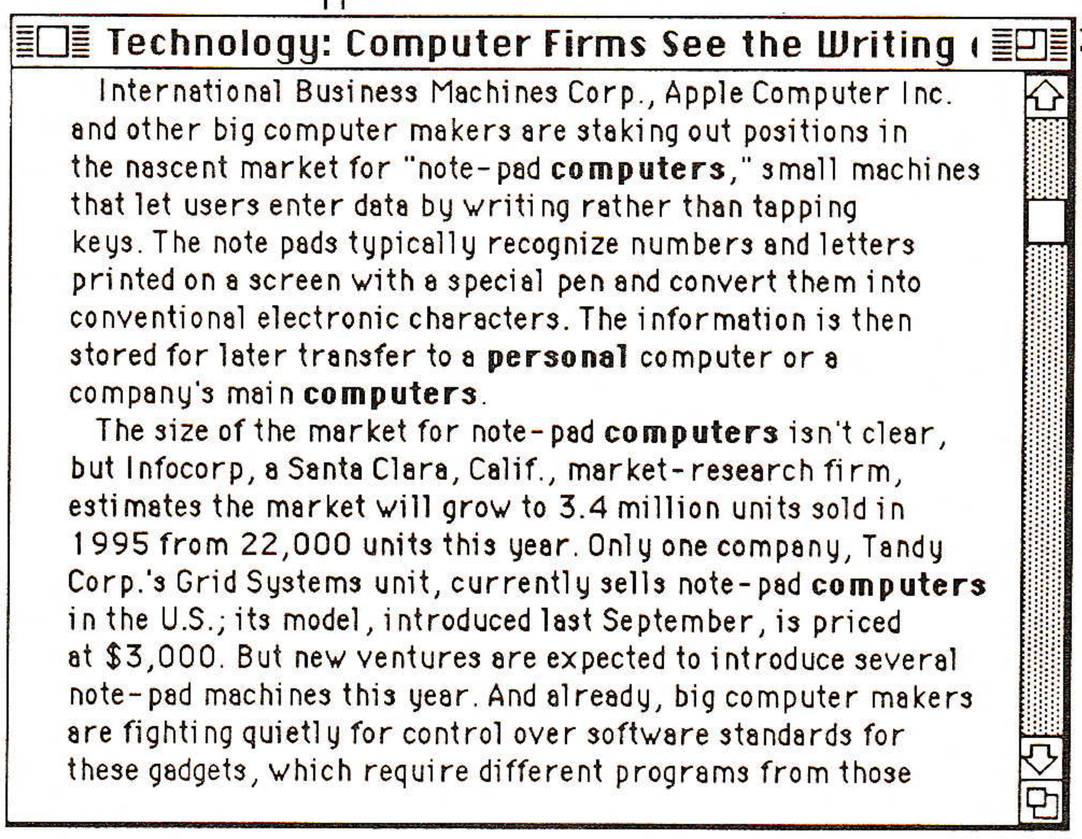
Step 1



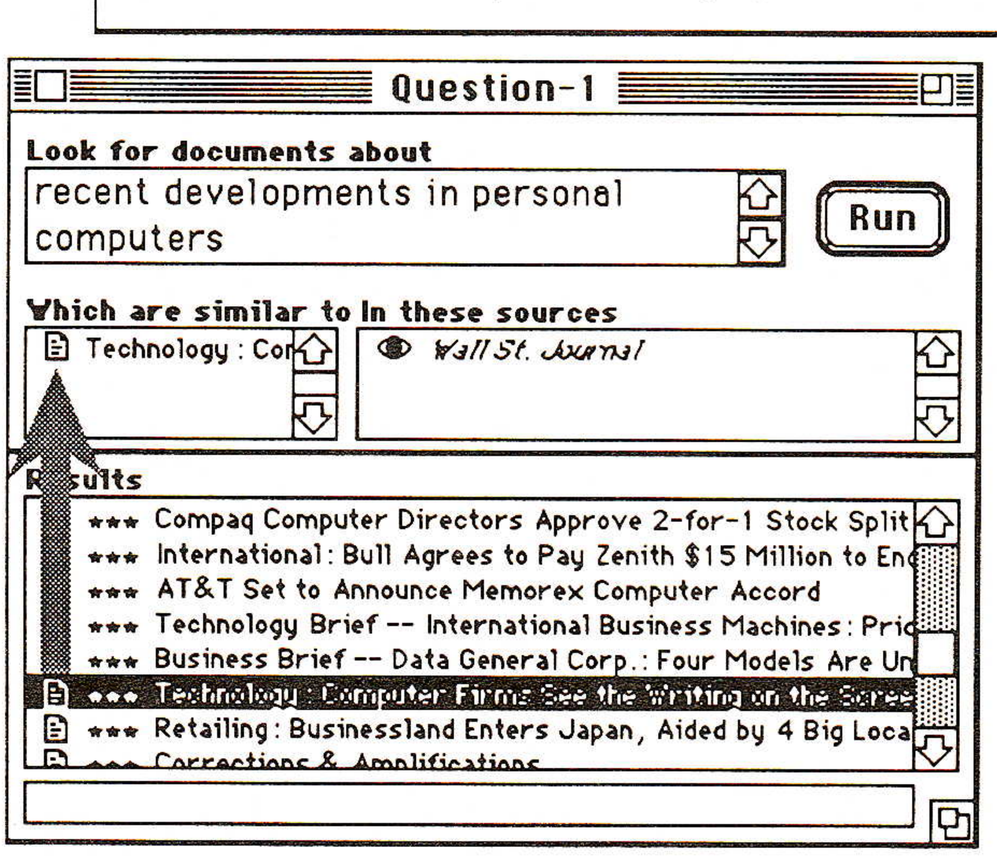
Step 2



Step 3



Step 4



continued on next page

Wide Area Information Servers *(continued)*

Contacting remote information sources

From the users point of view, a server is a source of information. It can be located anywhere that one's workstation has access to: on the local machine, on a network, or on the other side of a modem. The user's workstation keeps track of a variety of information about each server. The public information about a server includes how to contact it, a description of the contents, and the cost. In addition, individual users maintain certain private information about the servers they use. Users need to budget the money they are willing to spend on information from particular servers, they need to know how often and when each server is contacted, and they need to assess the relative usefulness of each server. This information helps guide the workstation in making cost effective decisions in contacting servers.

With most current retrieval systems, complications develop as soon as one begins dealing with more than one source of information. The most common problem is that of asking a particular question. For example, one contacts the first source, asks it for information on some topic, contacts the next source, asks it the same questions (most likely using a different query language, a different style of interface, a different system of billing), contacts the next source, and so on. One of the primary motivations behind the initial development of the WAIS system was to replace all this with a single interface.

With WAIS, the user selects a set of sources to query for information, and then formulates a question. When the question is run, the system automatically asks all the servers for the required information with no further interaction necessary by the user. The documents returned are sorted and consolidated in a single place, to be easily manipulated by the user. The user has transparent access to a multitude of local and remote databases.

A personal newspaper

In addition to providing interactive access to a vast quantity of information, the WAIS system can also be used as a rudimentary personal newspaper. A virtually unlimited number of queries can be saved, and updated at periodic intervals. To do this, the user's workstation is directed to contact each server at certain set times. When a source of information is contacted, any questions referencing that source are updated with new documents. The users can then easily browse through the results the next morning.

To make the ideal electronic personal newspaper, a system designer would need certain technologies which are not available today. Most computer screens are too small to allow efficient browsing of large amounts of text. Additionally, current data transmission speeds do not allow fast enough scanning if the text is not resident on the user's machine.

Despite current limitations, the WAIS system employs a number of features which will be found in the personal newspaper of the future:

- Clear displays of which questions have new documents
- Searches performed at night to hide communications delays
- Documents stored on disk for future reference
- Tools provided to quickly view stored documents

With these techniques, we have established a foundation of user support and acceptance.

Servers

The WAIS system was designed to be used by those who wish to sell information, as well as those who want to buy it. It provides a straightforward mechanism for indexing large amounts of data, making it available, and advertising the availability.

The system is flexible enough to provide for a variety of billing methods. A small database maintainer might make the information available through a telephone connection. Using a 900 number, the billing would be taken care of by the phone company. A slightly more sophisticated site might have a password and credit card billing system. High volume servers might want to set up flat fee contracts with customers. Other methods will certainly emerge as use increases. The system was designed to be as adaptable as possible to future financial arrangements.

As the dissemination of information becomes easier, questions of ownership, copyright, and theft of data must be addressed. These issues confront the entire information processing field, and are particularly acute here. The WAIS system is designed to keep control of the data in the hands of the servers. A server can choose to whom and when the data should be given. Documents are distributed with an explicit copyright disposition in their internal format. This is not to say that theft cannot occur, but if a client starts to resell another's data, standard copyright laws can be invoked.

Directory of Servers

As the WAIS system develops, sources of information will proliferate, making it impossible for any user to keep track of all servers that may be available at any one time. To help solve this problem, Thinking Machines is maintaining a *Directory of Servers* in a widely accessible location. The Directory of Servers contains indexed textual descriptions of all known servers. It is queried just like any other source. Instead of text documents, however, it returns source structures, specially formatted files which can be plugged into a question and used for queries.

For example, suppose you needed information concerning the current gross national product of Mali, but had no idea where to find it. You might first ask the directory of servers for "information about the current economic condition of Mali." The directory would return several documents, among them might be a source for the *World Factbook,* an on-line almanac maintained by the CIA. You would then use this document as the source field of a question, and re-run the query. This time, the system would contact the almanac, ask for the information, and return a document with the data you need.

Additionally, the Directory of Servers provides a means for information providers to advertise the availability of their data. When a new source becomes available, the developers can submit a textual description, along with the necessary information for contacting the server. This information is added to the directory, and becomes available to the public.

A common protocol for information retrieval

One of the most far reaching aspects of this project is the development of an open protocol. The four companies have jointly specified a standard protocol for information retrieval. Creating a market where new servers can be readily established requires an open, publicly available protocol. Ideally this protocol would be internationally standardized, yet flexible enough to adapt to new ideas and technologies; functioning over any electronic network, from the highest speed optical connections to phone lines.

continued on next page

Wide Area Information Servers *(continued)*

The use of an open and versatile protocol fosters hardware independence. This not only provides for a much wider base of users, it allows the system to seamlessly evolve over time as hardware technology progresses. It provides incentive to produce the best components possible.

For example, the protocol provides for the transmission of audio and video as well as text, even though at present most workstations are unable to handle them. However, they are free to ignore pictures and sound returned in response to questions, and to display and retrieve only text. This inability, though, does not hinder higher-end platforms from exploiting their greater processing power and network bandwidth.

The WAIS protocol is an extension of the existing *Z39.50* standard from NISO [3]. It has been augmented where necessary to incorporate many of the needs of a full-text information retrieval system [4]. To allow future flexibility, the standard does not restrict the query language or the data format of the information to be retrieved. Nonetheless, a query convention has been established for the existing servers and clients. The resulting WAIS Protocol is general enough to be implemented on a variety of communications systems.

The success of a WAIS-like system depends on a critical mass of users and information services. In order to encourage development and use, Thinking Machines is not only publishing a specification for the protocol, but is also making the source code for a WAIS Protocol implementation freely available. While this software is available at no cost, it comes with no support. We hope that it will facilitate others in developing servers and clients.

Future

In developing the WAIS system, the participating companies have demonstrated that current hardware technology can be effectively used to provide sophisticated information retrieval services to novice end-users. How this might effect information providers is not yet completely understood. The users at Peat Marwick found the technology useful for day-to-day tasks such as researching potential new accounts and finding resources within their own organization. Since these tasks are not restricted to the accounting and management consulting industries, we are optimistic that this type of technology can be fruitful and productive in many corporate settings.

The future of this system, and others like it, depends upon finding appropriate niches in the electronic publishing domain. Potential uses include making current online services more easily accessible to end-users; or allowing large corporations to access their own internal word processor files more efficiently. It is also possible that near-term development will focus on a single professional field such as patent law or medical research.

Summary

A unique alliance of four companies with complementary interests in the field of information retrieval have jointly developed a prototype which gives versatile access to full-text documents. The system allows users to retrieve personal, corporate, and wide area information through one easy-to-use interface. The WAIS project has shown that current technologies can be used to make useful, profitable, and convenient wide area information systems. The success of the project has convinced us that a WAIS-like system can be a valuable tool for corporate information retrieval.

Acknowledgements

The design and development of the WAIS Project has been a collective effort, with contributions and ideas coming from many people. Among them:

Apple Computer: Charlie Bedard, David Casseras, Steve Cisler, Tom Erickson, Ruth Ridder, Eric Roth, John Thompson-Rohrlich, Kevin Tiene, Gitta Soloman, Oliver Steele, Janet Vratny-Watts. *Dow Jones News/Retrieval:* Clare Hart, Rod Wang, Roland Laird. *Thinking Machines:* Dan Aronson, Franklin Davis, Jonathan Goldman, Chris Madsen, Harry Morris, Patrick Bray, Danny Hillis, Gary Rancourt, Tracy Shen, Craig Stanfill, Steve Swartz, Ephraim Vishniac, David Waltz. *KPMG Peat Marwick:* Chris Arbogast, Mark Malone, Tom McDonough, Robin Palmer. *Scolex Information Systems:* Art Medlar. Thanks also to Advanced Software Concepts for *TCPack* software.

For more information

Brewster Kahle
Thinking Machines Corporation
1010 El Camino Real, Suite 310
Menlo Park, CA 94025
415-329-9300 x228
brewster@Think.com

Thinking Machines Corporation
245 First Street
Cambridge, MA 02142
617-234-1000

References

[1] Salton, Gerald; McGill, Michael. *Introduction to Modern Information Retrieval,* McGraw-Hill, 1983.

[2] *DowQuest* promotional literature available from Dow Jones & Co. Inc., 200 Liberty Street, New York, NY 10281.

[3] "Z39.50-1988: Information Retrieval Service Definition and Protocol Specification for Library Applications," National Information Standards Organization (Z39), P.O. Box 1056, Bethesda, MD 20817. (301) 975-2814. Available from Document Center, Belmont, CA. Telephone 415-591-7600.

[4] Franklin Davis et al., "WAIS Interface Protocol Prototype Functional Specification," Thinking Machines, Corp. Available from Franklin Davis (fad@think.com) or Brewster Kahle (brewster@think.com).

[5] Barron, Billy, "Another use of the Internet: Libraries Online Catalogs," *ConneXions,* Volume 5, No. 7, July 1991.

[6] Quarterman, John, "Networks: From Technology to Community," *ConneXions,* Volume 5, No. 7, July 1991.

[7] Schwartz, Michael F., "Resource Discovery and Related Research at the University of Colorado," *ConneXions,* Volume 5, No. 5, May 1991.

BREWSTER KAHLE is Project Leader for the Wide Area Information Servers project. With Thinking Machines since the company was founded in 1983, Brewster architected the CPU of the Connection Machine Model 2 and lead the design of all of the custom chips. For the last 2 years he has been working on making the supercomputers a smart information server in a joint project with Apple, Dow Jones and Peat Marwick. He can be reached as brewster@Think.com.

Scolex founder **ART MEDLAR** has been active in the Information Retrieval field for many years. He designed an early WAIS prototype on the Apple Macintosh, and supplied the telecommunications code for the current system. His company provides IR and telecommunications consultation services to clients throughout the San Francisco Bay area.

[Ed.: This article also appeared in the September 1991 issue of *ONLINE Magazine.*]

Multimedia Mail From the Bottom Up
or
Teaching Dumb Mailers to Sing

by Nathaniel S. Borenstein, Bellcore

Abstract

Multimedia mail systems have exhibited great potential, but the widespread use of multimedia mail has so far been inhibited by the lack of interchange standards and the heterogeneity of mail-reading software. This article describes a new approach that seeks to break the existing log-jam and make multimedia mail a practical reality. The article begins with a brief summary of the state of the art in multimedia mail systems. It then outlines the new, "bottom-up" approach, and describes the configuration mechanism that is central to its operation. The article ends by outlining a vision of a new and better "lowest common denominator" for electronic mail.

The promise of multimedia mail

Electronic mail (e-mail) is a widely-used and much-appreciated technology. Ever since the inception of electronic mail, there has been much discussion of its even greater potential. For most people, e-mail today is a text-only medium, in which unformatted textual messages can be sent rapidly to even the most distant of correspondents. In principle, the limitation to plain text is artificial. E-mail is fundamentally capable of carrying richly formatted text, images, audio, video, and indeed anything that can be encoded in a digital form. In practice, however, the vast majority of the world's e-mail users are still restricted to plain text, due to a lack of interchange standards and a profusion of heterogeneous software for reading mail. The relatively few users of advanced multimedia mail systems such as *The Andrew Message System* [1] and *Diamond* [5] can only interchange multimedia mail with other users of the same software. An *Andrew* user and a *Diamond* user cannot, for example, send mail with pictures to each other. The result is that no multimedia mail technology has reached "critical mass" and made anything beyond plain text a part of the standard e-mail infrastructure for the masses.

The approach taken by most multimedia mail system to date can be characterized as a "top-down" approach. The developers of such systems said to their potential users, like Moses coming down from Mount Sinai, "Behold! I give you multimedia mail. All you need to do, in order to reap its blessings, is to change your mail reading program, your mail sending program, your text editor, your drawing editor, and generally everything about the way you work on a computer. Oh, and all your correspondents must do the same." When viewed in this way, it is perhaps not surprising that the world has not rushed headlong to embrace any of these systems.

The situation is best illustrated by considering the two different types of sites where *Andrew* is in use. At some sites, including the Carnegie Mellon University campus, where *Andrew* was developed, its use is nearly ubiquitous. (This was typically accomplished by administrative fiat.) Given this fact, the sender of a message can rely on the ability of the recipients to see a multimedia message in all its splendor. In such environments, a substantial portion of all mail messages contain at least multi-font text, and mail containing images, hypertext links, or other multimedia objects is not uncommon. At the other extreme, however, are sites where only a few individuals have elected to use *Andrew*.

While such individuals, like the users of any mail-reading software, may wax lyrical at times about the virtues of *Andrew,* they rarely, in practice, make use of its multimedia facilities, for the simple reason that their ability to send multimedia messages is useless if the people they're sending them to can't read them. Somewhere between these two situations, it seems, a community reaches critical mass with respect to the use of multimedia facilities. Clearly the Internet community as a whole is nowhere near reaching such critical mass, nor does it even seem to be moving in that direction.

It is difficult to doubt that multimedia mail would be greatly appreciated if it were widely available. The question, then, is how a transition can be effected from the current text-only mail world to a world of multimedia mail. The top-down approach that has been tried up to now shows little prospect of imminent widespread success. Convincing users to change to a new mail-reading program is, at best, a difficult proposition. It is made even more difficult by the fact that most users do not perceive themselves as "needing" multimedia mail and are unlikely to see its value until after they have already had it for a while.

Bottom–Up approach

What is needed, then, is a way to introduce multimedia mail without traumatizing users with an enormous transition, such as changing to a new mail-reading program. To put it starkly, what is really needed is to give the users of each existing mail reading program a new version of that program that has been enhanced to understand all the desirable kinds of multimedia mail.

When stated this way, the goal is nearly prohibitive. The cross product of the number of mail readers times the number of possible multimedia mail formats results in an enormous number of combinations. Moreover, each time the set of mail formats grows, each of the mail readers would need to be modified again. This is clearly impractical. However, there is a simplifying bottom-up architecture that makes the problem tractable once more.

In the bottom-up architecture, each existing mail reader is modified once, and only once. It is modified in a relatively simple way, without any knowledge about specific multimedia mail formats. In this modification, the only thing that changes is that, when the user asks to see a message, the mail reader first checks to see if the mail is non-textual (in Internet mail, this means checking the "Content-type" header field, as defined by [4]). If so, instead of simply showing the message body to the user, the mail reader checks a configuration file that lists a series of locally-recognized mail types, along with the locally-installed programs that can be used to view mail of these types.

The key point here is that each mail reader is modified only once, and that all mail readers are then able to obtain multimedia configuration information from a shared configuration file. Once this is the case, the addition of new media types at a site becomes a relatively straightforward matter: A binary program that can be used for viewing the type is installed, and a single line is added to the configuration file. Even if dozens of different mail readers are used at the site, their shared use of the configuration file means that users of any of those mail readers can now view the new type of mail.

Metamail

In the Bellcore prototype implementation, the software situation is simplified even further by the introduction of an intermediate program, called *metamail.*

continued on next page

Multimedia Mail From the Bottom Up *(continued)*

This program encapsulates all knowledge of the configuration files (called *mailcap* files in the prototype implementation), so that each mail reading program need only be modified to call *metamail* in order to display non-text mail. The resulting architecture is pictured graphically in Figure 1.

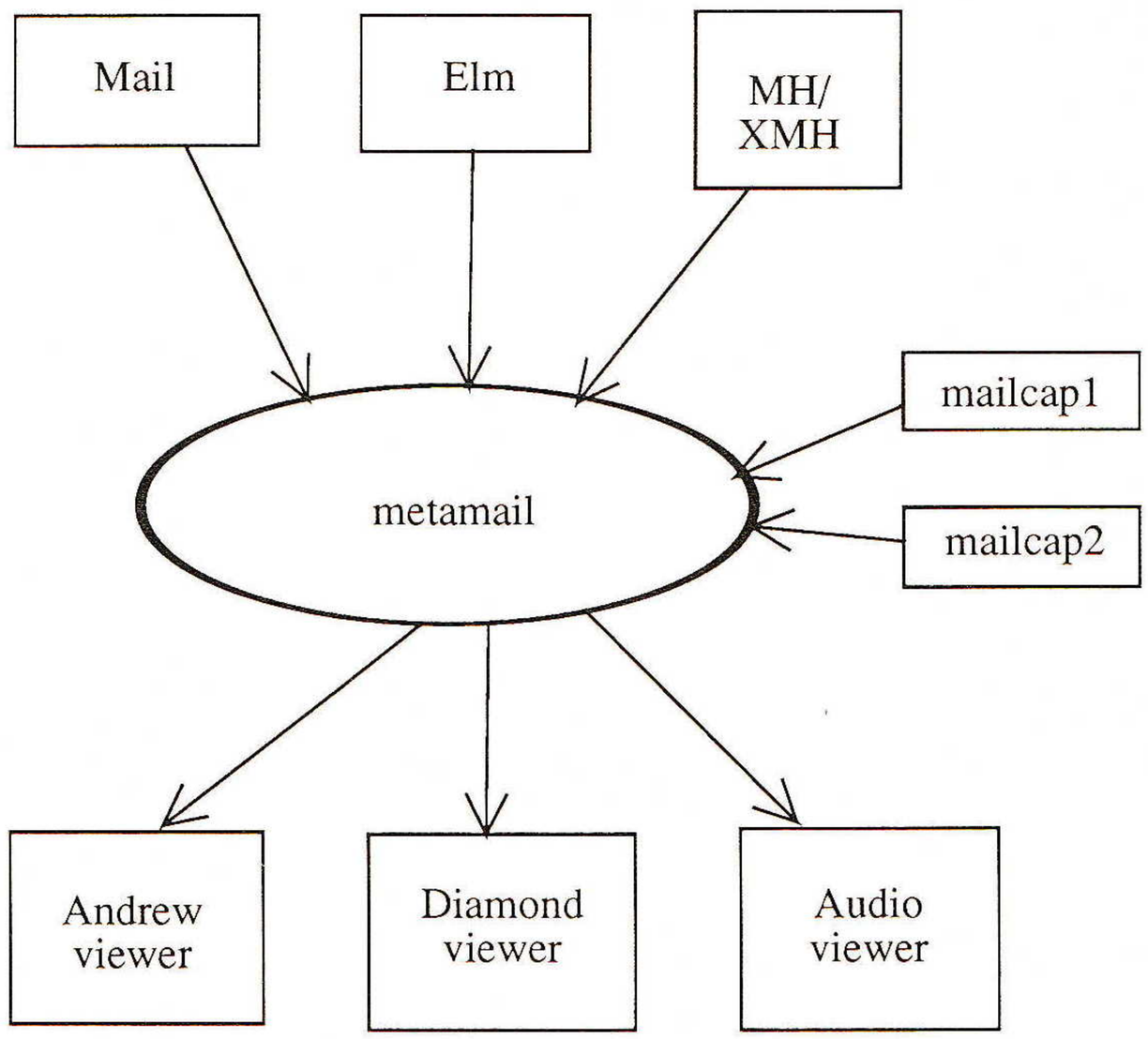


Figure 1: The *Metamail* "Bottom-Up" Architecture

Whether a mail reader includes knowledge about configuration files directly, or simply calls an external program like *metamail,* is not crucial. Some mail readers at a given site might work one way, and some the other way. What is more crucial, however, is that all the mail readers share a single configuration file mechanism, so that all the mail readers at a given site can be extended to handle new mail types via a common mechanism. Eventually, it is likely that users will gradually migrate to integrated mailers that handle multiple media types quite seamlessly, but the technique of modifying existing mail programs to be configurable for new media types is, at a minimum, extremely useful as a transition strategy for making multimedia mail widely available.

Figure 2 shows a user reading a message containing a picture, using one of the most primitive mail-readers available, Berkeley *mail*. This version of Berkeley *mail* includes a patch of approximately 30 lines that make it call the *metamail* program for non-text mail.

Mailcap files

The mechanism by which configuration information is conveyed to mail-reading programs (or to an intermediate program such as *metamail*) is the most critical part of the bottom-up approach. In order to permit multimedia mail to flourish in a heterogeneous environment, it is crucial that a wide range of mail reading programs should be able to share such a configuration mechanism. If a site administrator had to change a different configuration file, with a different syntax, for each mail reader at a site, it is unlikely that multimedia mail would ever work very well at sites that run a wide variety of mail reading programs.

However, if such a configuration mechanism is to be shared by all mail readers, it must be designed very carefully in order to insure that it provides enough information for a diverse range of mail-reading interfaces. The information that must be provided is not obvious without considering a range of mail readers.

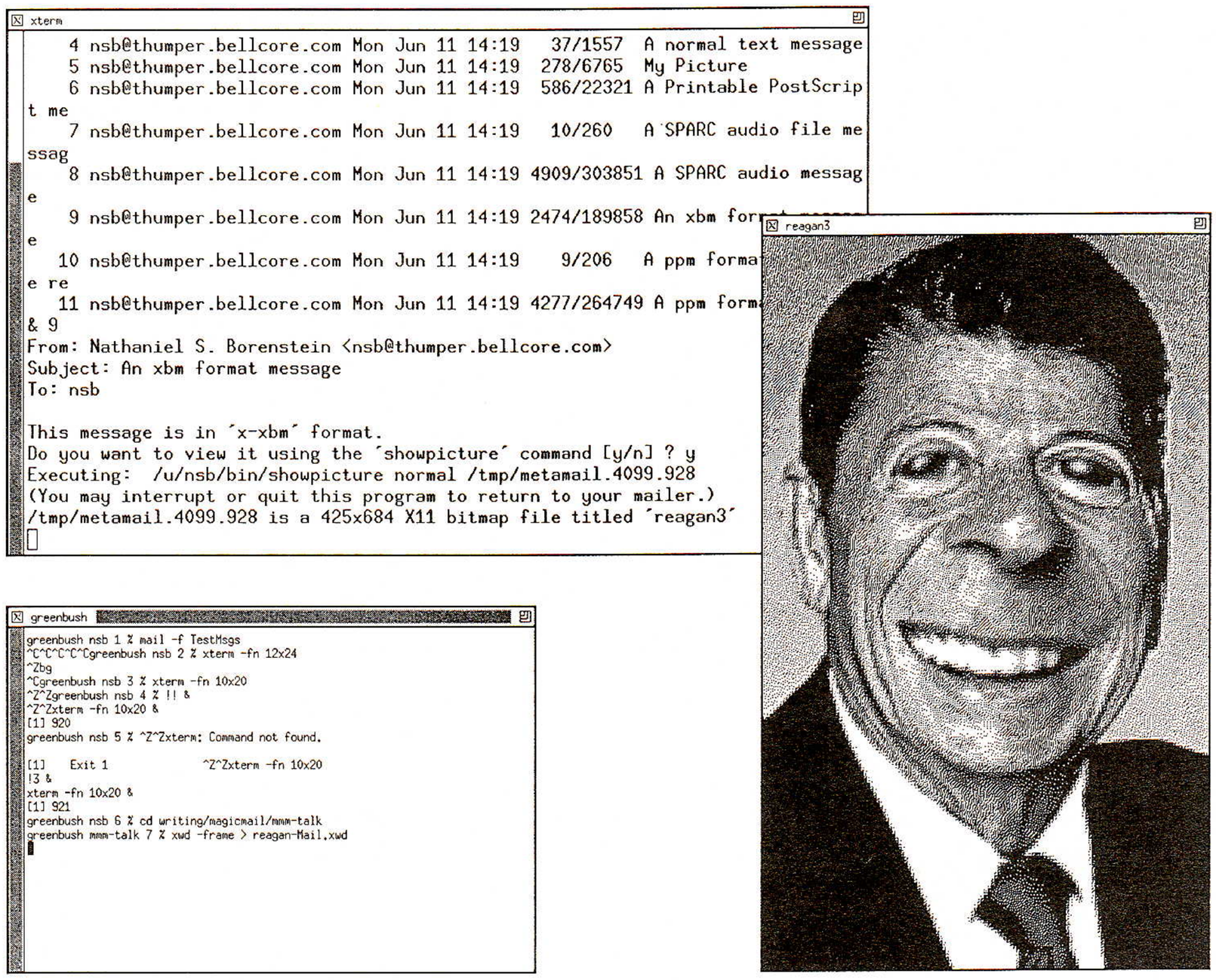


Figure 2: A Berkeley *mail* user reads a piece of multimedia mail

Program interaction

For example, a relatively "low-end" mail reader, such as the Berkeley *mail* program, never does anything more complicated than show text to the user. If the user sets an appropriate option, such text may be filtered through a paging program, such as the UNIX *more* program, in order to keep it from scrolling too quickly off the screen. If the *mail* program is configured to run an external program for some non-textual mail type, it wants to be able to tell that program to use a paging program if it is going to produce large quantities of output. However, it cannot simply assume that it is safe to send the output from such a program to a pager, because the program might instead want to interact with the user, conducting a dialog on the screen with which a paging program would substantially interfere. This might suggest that whether or not to run *more* or some other pager is a function of the external program, rather than of the *mail* program. This, too, is an oversimplification. Consider a window-oriented mail reader, such as *XMH, Andrew Messages, Xmail,* or *MailTool.* If an external viewing program produces large quantities of output when called from one of these programs, such output should not be passed through a pager, because it is being inserted directly into a scrollable window on the screen. On the other hand, if the external program needs to interact with the user on a terminal, a terminal emulator window needs to be created. In short, the situation is more complicated than it looks. The answer, in this particular case, seems to be that a pager is desirable only if it is appropriate for both the mail reading program and the external viewing program.

continued on next page

Multimedia Mail From the Bottom Up *(continued)*

The former information can be taken care of by the mail reading program (or, in the prototype implementation, by a command-line option to the *metamail* program), but the latter information must be encapsulated in the configuration file.

Format

The *mailcap* format used in the prototype implementation was the result of considerable trial and error, and the resolution of the kind of problems described above. A full specification of this format has been submitted as an Internet Draft, in order to promote a widely-shared format for the configuration file. Those interested in implementing a bottom-up mail reader, compatible with the ones described here, should consult the Internet Draft for a complete specification of the file format and location. In this article, we include only a partial description, to give the reader the flavor of the configuration format.

Configuration information is derived from a set of *mailcap* files, the location of which can be derived from a path given as the MAILCAPS environment variable, for which a standard default definition is also specified. Each *mailcap* file consists of comments (lines beginning with "#") and *mailcap* entries. Each *mailcap* entry (typically one line, although they can be continued on subsequent lines) describes how one particular type of multimedia mail can be handled. For example, consider this *mailcap* entry:

```
IMAGE/pbm; xloadimage -quiet -geometry +1+1 %s; nsb
```

This specifies that if a message has a header field of `Content-type: IMAGE/pbm` (the matching is case insensitive), then a file containing the body should be shown to the user with the `xloadimage` command, with the options specified. The "`nsb`" is a required field indicating the person who installed this *mailcap* entry locally. This particular *mailcap* entry is minimal, in the sense that it only uses the three fields that are required for each *mailcap* entry. However, additional fields are defined for specifying additional information about the format. For example, a `needsterminal` option specifies that a given application requires an interactive terminal, so that before it is called from a window-based mail reader, a terminal emulation window should be created:

```
Application/ATOMICMAIL; atomicmail %s; nsb ; needsterminal
```

Similarly, a `copiousoutput` option can be used to indicate that the application produces output that might be most appropriately passed through a pager such as *more,* depending on the windowing environment. Additional options can be used to specify external mechanisms to print messages, or to compose new messages of this type:

```
X-BE2; ezview %s; nsb; print=ezprint %s; compose = ez %s
```

The *mailcap* syntax is quite simple; the options relating to terminal characteristics are the most complex part. The syntax is fully specified in [2].

It should also be noted, in passing, that one other special header field must be recognized, along with `Content-type`. This field, `Content-Transfer-Encoding`, specifies how 8-bit or binary data is encoded for mail transport, since SMTP mail transport assumes 7-bit data of limited line length. The encoding mechanism is described in [4], and provides a standardized mechanism for encoding 8-bit and binary data for transmission via 7-bit SMTP mail.

The future of e-mail

Advanced multimedia mail systems such as *Andrew* and *Diamond* have shown the attractiveness and value of multimedia mail, but have for the most part failed to win over enough users to establish their high-level capabilities as part of the standard user's environment. More than most other computer applications, mail is inherently limited by the lowest common denominator. Unless nearly everyone with whom a user exchanges e-mail is able to properly handle advanced e-mail types, the user is unlikely ever to try to compose such types.

The real goal, then, for those who would have e-mail live up to its potential, is to create a new and higher-functionality lowest common denominator. A configurable bottom-up approach, such as the *metamail/mailcap* system described here, provides a transition path from the current world of text-only e-mail to a future in which the level of the lowest common denominator has been raised. But what will that raised level be?

It is unlikely, for example, that a new lowest common denominator could include full-motion video any time soon. Relatively few users have machines that are capable of displaying such data, and even fewer are connected by networks that can offer the requisite bandwidth. A more reasonable target, it would seem, for a new lowest common denominator would be a set of functionality that is accessible to nearly all users of modern computer system. As such a new lowest common denominator, I would propose the following four media types, along with auxiliary types such as the `multipart` type that allows these to be combined arbitrarily:

- *Text:* This is obviously already a reality. It seems plausible, in addition, to make a simple version of richly-formatted multifont text widely available, too. If the definition is simple enough, it will be a simple matter for a single-font terminal to remove the formatting information and show only the raw text. Thus a relatively portable version of formatted text could also become part of the lowest common denominator, if suitably standardized. Such a simple rich text format is defined in [4] and proposed as a standard facility for Internet mail. That document also proposes mechanisms to permit international text (text in multiple character sets) as a standard capability of Internet mail.

- *Image:* A growing percentage of computer users already work on computers with bitmap screens that are capable of displaying digital images. Moreover, nearly all such users are within shouting distance of a FAX machine. It is not unreasonable, then, to imagine that all computer users would have the capability to receive images in the mail; those without the necessary display technology should be able to specify the phone number of a FAX machine to which the image can be delivered.

- *Audio:* Similarly, more and more computers have audio capability, and users of computers that lack this capability are rarely far from a telephone, and could reasonably expect to have the audio portions of their messages delivered to the nearest telephone.

- *Computation:* Recent research by the author [3] has shown that it is possible to define a computer programming language that is both safe enough and portable enough to be executed automatically when received via insecure e-mail. Such programs, if defined in a suitably portable language, can run on any computer terminal in the world.

continued on next page

Multimedia Mail From the Bottom Up *(continued)*

Thus it is not unreasonable to imagine computation, in a suitably standardized language, as part of the new lowest common denominator, allowing users to send each other messages that interact directly with the recipients and take actions based on that interaction.

Crucial to the evolution of a new lowest common denominator is clear, concise, and implementable standards. A recent Internet memo [4] defines an interoperable set of mechanisms and formats that are intended to evolve into such standards, and that seek to define a new lowest common denominator for electronic mail. The bottom-up approach described in this article is wholly compatible with these mechanisms, though it is not the only possible way to implement them.

Acknowledgements

The development of *metamail* and *mailcap* was stimulated by an initial conversation with Jonathan Rosenberg. Along the way, I've had immense amounts of help, from more people than I can really recall. I'm particularly grateful to Steve Uhler for picking up the ball and running with it, to Mike Bianchi for words of support at just the right moment, and to Bob Kraut, Al Buzzard, and the many others at Bellcore who have been extremely supportive and helpful in this work.

References

[1] Borenstein, Nathaniel S., & Chris A. Thyberg, "Power, Ease of Use, and Cooperative Work in a Practical Multimedia Message System," *International Journal of Man–Machine Studies,* April 1991.

[2] Borenstein, Nathaniel S., "A User Agent Configuration Mechanism for Multimedia Mail Format Information," Internet Draft `borenstein-configmech-00`, June 1991.

[3] Borenstein, Nathaniel S., "Secure and Portable Active Messaging: A New Platform for Distributed Applications and Cooperative Work," submitted to *Communications of the ACM.*

[4] Borenstein, Nathaniel S., & Ned Freed, "Mechanisms for Specifying and Describing the Format of Internet Message Bodies," `Internet Draft 822ext-messagebodies-00`, June 1991.

[5] Forsdick, H. C., Thomas, R. H., Robertson, G. G., & Travers, V. M., "Initial Experience with Multimedia Documents in Diamond," Computer Message Service, Proceedings IFIP 6.5 Working Conference, IFIP, 1984.

(A longer version of this article is being submitted as a paper to the USENIX technical conference).

NATHANIEL S. BORENSTEIN received his B.A. (1980) in Mathematics and Religious Studies from Grinnell College, and his M.S. (1981) and Ph.D. (1985) in Computer Science from Carnegie Mellon University. He worked for over four years on CMU's *Andrew* project, where he was the author of much of *Andrew*'s end-user interface, and was one of the primary authors of the *Andrew Message System.* For the past two years, he has worked for Bellcore, where his research has focused on making multimedia mail more powerful and at the same time more easily integrated into existing mail environments. His second book, *Programming As If People Mattered: Friendly Programs, Software Engineering, and Other Noble Delusions*, has just been published by Princeton University Press. He can be reached via e-mail as: `nsb@thumper.bellcore.com`.

CD Review

The TCP/IP CD from SRI International. This is an ISO-9660 (High Sierra) format compact disk. Files are formatted for UNIX and MS-DOS systems and may be readable on other systems supporting ISO-9660 file systems. Priced at $395, it includes one free update disk. Call 415-859-NETS (That's 415-859-6387) or send electronic mail to `TCP-IP-CD@NISC.SRI.COM`. [See *ConneXions* Volume 5, No. 8, August 1991 for more information].

Organisation

This nifty little album was released earlier this year and contains a collection of tracks of some interest to readers of this journal. The works are collected together in an unusual "Concept Album" structure with track listings as follows:

- Doc:

FYI	For Your Information — telstar, long distance love, etc.
Humor	Dry Mr. Protocol and friends
IEN	Internet Engineering Notes
IRG	Internet Resource Guide (green)
Misc	A list of three letter acronyms (TLAs)
NETINFO	Useful things like an X.25 Spec.
Protocol	Hot gossip
RFC	Request For Comments — a kind of audience appeal
THEnet	YAG (Yet Another Guide)
Worm	Collected turnings of the worm — Gothic, man

- Mail:

Bind 85–91	Ramblings from over 5 years of life in the swamp
Domain 83–91	Even longer from the ancestral home
TCP-IP 82–91	A cure for all those cuts and bruises

- SRC:

BSD	Collected comms code from a commune somewhere west of Kansas
ISODE	Politically correct comms code
Mac	NCSA Telnet — need i say more?
PC	Turn a PC into something useful e.g., KA9Q
UNIX	Networked files, time, serial lines and management plus trivia

The conceptual approach seems to be not quite encyclopedic. For instance, the kitchen sink cannot be heard anywhere here, nor is an X Window opened upon any view. The sum of the parts is found, *IFIND,* to be somewhat less more than the whole. This may be caused by the curious 56 Kbps ARPANET 6-bit under-sampling compared with European standards.

Recommended

All in all, they're playing my song, man. I think this CD technology can be summed up in the phrases "what less can you ask for?" and "quite definitely up to scratch."

—Jon Crowcroft, University College London

Is Resource Discovery Hacking?
The Great Measurement Debacle

by Carl Malamud

Introduction

Mike Schwartz is an Assistant Professor at the University of Colorado. He specializes in an emerging area of network applications known as *resource discovery*. Instead of some purely academic pursuit such as reinventing the remote procedure call, Schwartz tries to conduct research on the existing Internet and how to use it more effectively.

Resource Discovery

There are several examples of resource discovery. Schwartz has developed software called *netfind,* which attempts to locate electronic mail addresses. Rather than using a single source of information, *netfind* uses *finger,* SMTP, DNS, and a variety of other sources to try to locate users. The Corporation for National Research Initiatives has developed a system similar to *netfind,* known as the *Knowbot Information Service* (KIS) and is working on intelligent search techniques for information archives.

The *Archie* project at McGill University compiles a list of files available via Anonymous FTP. *Archie* allows the user to discover where in the Internet a file exists. Other projects include *Prospero* developed by Clifford Neuman of ISI and *WAIS,* pioneered by Brewster Kahle of Thinking Machines.

Study

A recent series of incidents underscores many of the issues in this emerging area. Schwartz was working on a research project to measure the nature of connectivity to the Internet. The study is quite simple. Using a statistical sample of hosts on the Internet, Schwartz tries to see what services they offer.

This study measures, in a statistical fashion, the nature of connection to the Internet. As a longitudinal study, Schwartz hopes to see if the nature of that connection changes over time. For example, people might start disconnecting themselves because of security concerns, as Dr. David Clark of MIT has predicted.

Services

The study is interesting in several respects. First, it is not based on simulation or other theoretical models: it uses the real Internet. His study starts with a database of some four hundred thousand hosts in over 12,000 different domains. Then, Schwartz picks a sample of that population, doing on average 3.9 name lookups per domain and attempting 5.7 connections per domain. The study attempts to open the following ports:

Port Number	Service
13	daytime
15	netstat
21	FTP
23	Telnet
25	SMTP
53	DNS
79	Finger
111	portmap
513	login
540	uucp daemon
543	klogin
544	krcmd kshell

The services were chosen carefully to try to yield information about the type of connectivity a particular host offers.

For example, ports 543 and 544 are *Kerberos*-based remote access procedures and indicate the presence of a secure gateway. The services chosen were those that one can expect to be run on every machine in a domain, thus making a statistical analysis valid. For example, one could expect SMTP on every port, but specialized protocols such as Z39.50 would probably run only on selected servers.

Only TCP ports were picked since with UDP there is no application-independent way to see if a UDP-based service is running. With TCP, one can simply see if a connection attempt succeeds to discover if the service is running. Note that successfully opening a TCP connection does not mean that the service itself will be used. Most services have some form of access control, such as a password. The Schwartz study immediately disconnects the TCP connection and never establishes an association at the application level.

The program is quite careful not to overload either individual machines or the Internet. No more than 3 connection attempts are made on no more than 3 machines in a particular domain. The software uses 20 concurrent threads, ensuring that no more than 20 probes are active at any one time on the network. The program takes roughly a day to run and contributes, on the day it is run, an increase in Internet traffic of roughly 0.5%

Sound harmless enough? Small-scale versions of the study were run in August, 1990 and February, 1991 on the *finger* and SMTP ports as a way of determining the potential scope of the *netfind* tool. A few astute system managers noted the probes and contacted the *Computer Emergency Response Team* (CERT) and University of Colorado system managers, both of whom had been advised before each of the studies.

In August, 1991, Schwartz expanded the study to look at more services. He began running his study and, again, a few managers contacted the CERT and system managers at the University of Colorado to find out what was going on. In all cases, the resolution was pretty simple. Schwartz explained his study and people said "fine." However, because the study was bigger and there were more notifications to the CERT, Schwartz volunteered to advise people about his activities.

Notifications

Two methods were used to notify system managers. First, mail was sent to the postmasters in the domains affected. Mail started to get sent, but after 4,000 messages went out it was evident that less than half of the domains had a valid postmaster account. The remaining half of the messages were undeliverable, served by an automatic response, or otherwise did not succeed in reaching a postmaster.

Schwartz also posted messages on USENET to advise people of what he was doing. Schwartz posted a message on `alt.security` that began:

> "I am in the midst of conducting a series of experiments designed to measure changes in service level reachability in the global Internet, to help characterize the extent to which institutions are distancing themselves from the Internet..."

The note went on and explained the methodology, the fact that the CERT was aware of the activity, and offered to make more information available. Messages started coming back from a few paranoid system administrators threatening to put Schwartz in jail, to call the FBI, and to otherwise cause severe problems if he attempted to open a connection on any of their hosts.

continued on next page

Is Resource Discovery Hacking? *(continued)*

Strong reactions

A large computer company, for example, explained that they were a "commercial user on the Internet" and that any attempt to conduct the study would "be treated as an intrusion and will be considered a prosecutable offense which we will pursue."

One of the most vociferous messages came from a systems manager at a startup well-known for making PC clones, who explained to Schwartz:

> "If you were to traverse our gateway(s), I'd simply call the FBI. You're way out of line, and really asking for significant legal problems. I in *no way* support what you're starting to do, it's wrong and UNETHICAL TO ENTER OTHER SITES, FOR WHATEVER 'REASON' YOU'RE TRYING TO YOURSELF JUSTIFY IT/THIS FOR." [sic]

(An interesting note, by the way, is to look up this company in the Domain Name Service. One sees that they use UUCP and UUnet to receive their mail, are not part of the Internet, and would not have been hit by this study.)

Whew! That's enough to make most researchers sit up and take notice. Lawyers still cost money, after all. With the threats of law suits coming in, Schwartz decided to temporarily table the study.

Objections

The next step was to reevaluate if this study had somehow violated the boundaries of what one is allowed to do on the Internet. Several arguments were raised in objection to the study:

- The study was a waste of machine resources
- The study was a waste of Internet bandwidth
- The study was a waste of network manager bandwidth by making them track down the intrusion and see if it was legitimate
- The study was a security violation

The first two arguments are really red herrings. The amount of machine resources was extremely small, and the load on the Internet was not substantial. The study was carefully engineered to make sure that a particular name server was not hit too many times, that a particular domain was only hit a few times, and that, in the case of failure, the study would simply give up and move on rather than keep trying.

The last two arguments, however raise fundamental issues. First, there is the question of the network manager. Several astute network managers noticed repeated tries to connect to ports on their machines and suspected an automated breakin attempt was in progress. They spent time trying to determine if they were under attack.

One of the outcomes of the Schwartz study was a realization that there needs to be some way of making system managers aware of this type of work. The `alt.security` posting didn't appear to reach a wide audience. A full-fledged CERT alert would be inappropriate for this activity. Some form of background information distribution is needed for activities of this sort.

The fundamental question is whether or not this type of work is allowed. The Internet Measurement study is one of an emerging class of applications.

People are talking about navigating the network with a joystick, searching FTP resources for relevant information in a network-wide search, and a wide variety of other applications. Schwartz may be the first to try this particular massive *ping,* but he certainly will not be the last.

It is widely accepted that a single user can use the *finger* service to look for a single user on a single machine. Is it acceptable to methodically use *finger* to try and build a directory?

What about methodically using *finger* to conduct research? Do we make a differentiation between research and commercial use? Between statistical analysis and the preservation of individually identifiable data? Do we care what type of user is conducting the probe?

Issues

There are three issues that can be raised by this type of work:

- A definition of the security perimeter on the Internet
- Privacy constraints
- Scaling

Most would agree that you are allowed to FTP to a host and attempt to login with the username "anonymous." This is a service, if it is there, that is available to the general public. Attempting to use anonymous FTP, even if the service is refused, is not a violation of the security perimeter. On the other hand, it is generally accepted that trying other usernames at random would in fact be a violation of the security perimeter.

By that definition, what Schwartz was doing did not violate the security perimeter of the Internet. He simply attempted to open a so-called *well-known port.* He did not try to login or to conduct any sort of attack on the perimeter.

Internet reaction

The opinion of many members of the Internet community was solicited on this issue. Almost unanimously, IAB, IESG, and NSF officials responded in a similar manner: "So what's the problem?" A typical response was from Dr. David D. Clark of MIT:

> "In general, the community has considered it out of bounds to attack systems, even with no malice, to see if they have a security flaw. We consider it in bounds to touch a system (e.g., *finger*) to see if it is there. So, without reading his info files, I would conclude that this experiment is within the bounds of acceptable behavior."

If a service is available to the public, such as the *finger* service, one is able to use the service. By this analysis, attempting to open TCP ports is perfectly legitimate. Trying to hack a username/password is not.

If you have a well-known port on the Internet, you should be prepared for people to attempt a connection. Using this analysis, the Schwartz study was perfectly legitimate. The activity he was conducting would have been perfectly legitimate if done manually and the fact that a program was conducting the activity in a wholesale fashion does not change the security issues.

Secure gateways

Many institutions have begun using secure Internet gateways. These gateways protect the hosts inside of the domain from exactly this type of intrusion. In other words, organizations that don't want to be subject to research like that of Mike Schwartz ought to install a secure Internet gateway. This is exactly what most major computer companies, including Sun, DEC, and IBM, have done.

continued on next page

Is Resource Discovery Hacking? *(continued)*

Privacy

Privacy falls within the same boundaries. Organizations that wish to protect the privacy of their users turn off services that violate that privacy. For example, many sites choose not to offer the *finger* service to the outside world. The Schwartz study would, in this case, attempt to open the *finger* port and record a failure, indicating that this particular site is not offering the *finger* service. If three hosts in a domain all fail to offer the service, the study then assumes that the service is not offered in the domain.

It is important to realize that we need both protected and unprotected forms of connection to the Internet. The Internet is a global community and simply prohibiting people from walking the streets of the global village is not an adequate solution. If you don't want people looking in your window, then pull the shades.

There is an aspect of privacy that does need to be considered, however. Some information, based on a single query, might not violate privacy. For example, given a real name, we can get an electronic mail address. However, one nature of a computer network is that we can make repeated narrow queries to amass large quantities of information. This is one of the fears of offering X.500 services: some people might make use X.500 to make copies of an entire directory for marketing or other purposes rather than finding narrow pieces of information.

Scaling

The Schwartz study does raise an important question when it comes to scaling: even if we allow a single Schwartz to conduct research what happens when 1,000 high school students begin emulating him? One person *pinging* a port is not a problem: millions of people doing so certainly is. Another example of this issue is the *Archie* project. One *Archie* server getting a directory listing off your system is not a drain on resources, but one million *Archies* doing so would be.

We need to support resource discovery on the Internet, but we also need to think very carefully about setting up an infrastructure to support this class of applications. This is a crucial area of research. A single centralized directory will not solve people's desire for information. Different users need different kinds of information and a resource discovery infrastructure needs to support a wide variety of different classes of modules, ranging from simple IP *pings* to Knowbots.

How to solve this type of problem is currently under study by an Internet Research Task Force research group headed by Schwartz. The research group is focusing on the technical issues involved in resource discovery. The IETF has also made a few stabs at resource discovery, but it is evident that we need to know a lot more about this class of applications before any hit the RFC stage.

Conclusion

The need to learn more is one of the most compelling reasons to allow Schwartz to move forward on this type of application. If qualified researchers are not able to expand the functionality of the Internet, we will cease to make progress. Only by seeing what happens when these types of experiments are being conducted can we begin to think about an Internet-wide infrastructure to support resource discovery.

For further reading

- Papers written by Mike Schwartz can be obtained by anonymous FTP from `latour.cs.colorado.edu`. See also the article by Schwartz in the May 1991 issue of *ConneXions* (Volume 5, No. 3).
- *Knowbots* are described in Malamud, *STACKS—The INTEROP Book,* (Prentice Hall, 1991). To access the Knowbot Information Service, send mail to `KIS@NRI.Reston.Va.US` and put "?" in the first line of the message.
- WAIS is described in this issue of *ConneXions,* see page 2–9.
- *Archie* can be accessed by Telnet at `Quiche.CS.McGill.CA`. Login as "Archie," no password is required.
- *Prospero* can be retrieved by anonymous FTP from Internet host `cs.washington.edu`.

CARL MALAMUD (`carl@malamud.com`) works with Mike Schwartz on issues of resource discovery. Malamud is the author of several books, including *STACKS—The INTEROP Book.* He is currently traveling around the world doing research for a technical travelogue.

The Internet Gopher: A Distributed Information Service

What is it?

The *Internet Gopher* is a distributed document delivery service. It allows a neophyte user to access various types of data residing on multiple hosts in a seamless fashion. This is accomplished by presenting the user a hierarchical arrangement of documents and by using a client-server communications model. The Internet Gopher Curses Client allows a user on a terminal to access the vast array of information available on various gopher servers. The Internet Gopher Server accepts simple queries, and responds by sending the client a document.

Software

Also included in the release are experimental clients and servers for real-time radio, utilities for using gopher in shell scripts (written in *perl*), and some sound utilities for NeXT machines. Other Internet Gopher Software available includes:

- Macintosh Gopher client written in *HyperCard.*
- Macintosh Gopher Server software.
- PC Gopher Client with a Borland *Turbo Vision* interface.
- Full Text Indexing servers for NeXT machines.
- NeXT Gopher client (provided by Max Tardiveau of the University of St. Thomas.)

Getting a copy

All of this software is available for anonymous FTP from `boombox.micro.umn.edu` (128.101.95.95) in the `/pub/gopher` directory.

The Internet Gopher Development Team can be reached via e-mail as `gopher@boombox.micro.umn.edu`. If you prefer paper we can be reached at:

Internet Gopher Team
Computer & Information Services
University of Minnesota
132 Shepherd Labs
100 Union Street SE
Minneapolis, MN 55455

Announcement and Call for Papers

The *3rd Joint European Networking Conference*—"Building Research Networking for Europe" will be held in Innsbruck, Austria, May 11–14, 1992. The conference is organized by RARE (Reseaux Associes pour la Recherche Europeenne) in cooperation with ACM SIGCOMM, ACOnet, EARN, EurOpen, and the Internet Society.

Goals

This conference aims at informing the participants about the state of the art in networking and about building new and better network services. It will provide a forum for the presentation and discussion of technical and strategic topics related to the provision of networking services for research and higher education, as well as corresponding research and development activities. As a result participants will get an understanding of the issues in European Networking.

The conference addresses technical, managerial and end-user support staff from local, national and international service providers as well as application developers, policy makers and representatives of funding bodies, advanced user groups and standards organizations.

Much emphasis will be placed on cooperation between networking services. The conference will continue and enhance the discussion between members of different networking communities, building on the positive experiences of the earlier Killarney and Blois Conferences. This conference is *the* forum on networking in Europe and presents a unique opportunity to meet key people active in networking today.

Venue

The city of Innsbruck is situated in the middle of Tirol, one of the most famous holiday areas in Austria. Innsbruck offers history and traditions as well as an up-to-date infrastructure. The conference will take place in the Innsbruck Convention Centre—Kongresshaus Innsbruck—a spacious centre of high international standard.

Invitation and preliminary programme

A first invitation and preliminary programme, including information on how to register will be distributed in January 1992. Please contact the RARE Secretariat if you want to make sure you receive the invitation.

Call for papers

As for the past Joint Networking Conferences, the programme will include a combination of solicited and submitted papers. Presentations taking 20, 30 or 45 minutes are invited. One-page summaries of proposed papers should arrive at the programme chair not later than November 17, mentioning the topic against which the author would like his paper assessed and the expected length of the conference presentation. It is again intended to publish proceedings of the conference in *Computer Networks and ISDN Systems*. Full papers must be provided by the start of the conference at the latest.

Topics

Topics for submitted papers should be related to the major headings of the general outline of the conference given below:

Users, User Support & Group Communications

- User Support
- User View of European Networked Resources—User Requirement
- Impact of Networking
- Teleconferencing
- Videoconferencing

Infrastructure: Coordination & Management

- Multiprotocol Backbone Infrastructure
- Network Management
- Operational & Interworking
- Quality of Service—Concepts, Performance, Measurement
- Security Implementation and Operation
- Services Management
- Gateways

Coordination of Applications & Services / Projects

- Access Control and Authentication
- Directory Services
- Distributed Computing
- Distributed Services Management
- Documentation Format, e.g., ODA
- File Servers
- Information Services (Library etc.)
- Message Handling Systems (MHS)
- Naming and Address Management
- High Performance Computing
- Visualisation

Technology

- New Technology (ATM, Frame Relay)
- Products
- Security Techniques

Policy, Funding & Futures

- Economic Impact of Networking
- Electronic Publishing & Intellectual Property Rights
- International Export—Legal Restrictions
- Is there life after COSINE?
- Security Policy

International Success Stories—Advanced Uses / Users

- Collaborative Research
- Distance Learning

Status reports of national initiatives and European projects

- COSINE
- Country Reports
- RARE Working Groups
- Standards
- EARN/EurOpen/RIPE

continued on next page

Announcement and Call for Papers *(continued)*

Call for posters and demonstrations

As in previous years, a poster wall will be available for the display of posters. Participants are invited to submit a poster presentation of the project they work on or a topic of common interest to the conference participants. The programme committee will select the two best posters during the conference for inclusion in the conference proceedings.

During the conference there will be the opportunity for participants to present their project or activities in the form of a demonstration, either as part of their presentation or separately. Requests for demonstrations should be made through the RARE Secretariat, specifying technical requirements. X.25 and IP connectivity will be provided.

Further information

For further information contact:

RARE Secretariat
P.O. Box 41882
NL-1009 DB Amsterdam
The Netherlands
Tel: +31 20 592 5078
Fax: +31 20 592 5043
E-mail: raresec@nikhef.nl
X.400: C=nl; ADMD=400net; PRMD=surf; O=nikhef; S=raresec

Christian Michau
Programme chair
Tel: +33 1 44274260
Fax: +33 1 44274261
E-mail: michau@frors12.bitnet

Program Committee

Christian Michau, Vint Cerf, Howard Davies, Marieke Dekker, Jill Foster, Frode Greisen, Krzysztof Heller, Dennis Jennings, Barry Leiner, Manfred Paul, Paul-Andre Pays, Paul Van Binst.

Coming in future *ConneXions*

We have many exciting articles "in the pipe." Some highlights include:

- Components of OSI: ASN.1
- Components of OSI: X.25—The Programmers Perspective
- The Xpress Transfer Protocol (XTP)
- An overview of OSI NSAP Addressing in the Internet
- GOSIP Challenges in the DoD
- ITU Adopts a New Meta-Standard: Open Access
- INTEROP 91 Fall in words and pictures

So make sure your subscription is current. For any questions contact:

ConneXions—The Interoperability Report
480 San Antonio Road, Suite 100
Mountain View, CA 94040–1219
USA
Phone: +1 415-941-3399 or 1-800-INTEROP (Toll-free in the USA)
Fax: +1 415-949-1779
E-mail: connexions@interop.com

New Internet Technology Series Available

Background

More than 1,200 RFCs have been issued in the last twenty or so years. For those knowledgeable in networking, each new RFC is warmly greeted and slid effortlessly into its proper niche in the grand scheme of Internet information. However, for those just introduced to the Internet, the plethora of information contained in the Request For Comments documents can be daunting. What is a protocol? Which RFCs are protocols? Which protocols are standards? Which standard protocols are required? What new protocols are being tested? Isn't there any general information? And what is this OSI stuff anyway?

Both to help neophytes snatch order from the jaws of confusion and to aid old hands with sorting the vast amount of information, the Network Information Systems Center at SRI International, under the editorial guidance of Dr. Vinton Cerf, has developed the *Internet Technology Series* (ITS). The ITS has three parts: the Internet Technology Handbook, the TCP/IP CD, and a subscription service.

Internet Technology Handbook

The *Internet Technology Handbook* is a 5,000 page hardcopy collection of RFCs gathered into six volumes. Based on the information discussed in RFC 1200, "IAB Official Protocol Standards," the Handbook organizes a core set of RFCs into ten sections. Each section briefly explains a specific networking topic, such as the internet layer, the transport layer, routing, network management, or applications, and presents RFCs that relate to that area. Other sections provide information on more general topics, such as Internet policies and architectural models. Many FYI RFCs are also included, presenting introductory Internet information. The Internet Technology Handbook updates and obsoletes the popular DDN Protocol Handbook previously compiled by SRI.

TCP/IP CD

The *TCP/IP CD* contains all the RFCs that are currently available online. This amounts to more than 500 RFCs that users can now access locally, thanks to the latest in CD technology. All online IENs are also included. Plus, archives of the *TCP-IP* and *Namedroppers* (domain naming) mailing lists are provided. An easy-to-use search program called *IFIND* allows users to specify and locate needed information contained in these files. In order to take advantage of the capacity of the CD, several other public domain files and applications have been included as a value-added bonus. One update to the CD is included as part of the order. (See review on page 17).

Internet Technology Subscription

The contents of the Internet Technology Handbook will be updated for one year by means of the *Internet Technology Subscription* service. This service provides hardcopies of RFCs that pertain to the Handbook, along with a summary of their contents and guidelines explaining their relevance. The subscription service ensures that the Internet Technology Handbook will never go out of date. After one year, the subscription service can be ordered independently.

More information

The Internet Technology series is available as a package, or each component can be ordered separately. Of course, SRI continues to be an online repository for RFC documents, making them available for FTP from the host `FTP.NISC.SRI.COM`. They are also accessible via electronic mail via a message to `mail-server@nisc.sri.com` with "send rfcnnnn" in the body of the message (where "nnnn" is the number of the RFC). Users can order paper copies of the RFCs individually as well. For further information, send a message to `nisc@nisc.sri.com` or call 1-415-859-NETS (That's 415-859-6387) or 1-415-859-3695.

Book Review

STACKS: Interoperability in Today's Computer Networks, by Carl Malamud, Prentice Hall, ISBN 0-13-484080-1.

[*STACKS* is also known as "The INTEROP Book"—a new source of information provided by the INTEROP Conference and Exhibition—and was produced as a professional reference book in cooperation with Prentice Hall. *ConneXions,* in which this review appears, is also a publication of Interop, Inc. —Ed.]

Organization

STACKS is an interesting meeting of the technology, politics, and usage of interoperable systems. In 285 pages, Malamud runs the gamut from new interpretations of the OSI model, LANs and WANs, protocol stacks, environments for distributed computing, network security and management, projects in high-speed networking, and how to find things in our current and future collection of interconnected networks.

As you might guess from the editor's note above, *STACKS* was conceived as a means for putting into perspective the entire INTEROP Conference and Exhibition. It succeeds marvelously at this. Malamud's style is seemingly both structured yet informal. For each area, he explains the origins and core aspects of the technology, and usually presents a case study describing its usage in the real world. His style is also disarmingly informal, as he easily glides from one topic to another in trying to make sense of the big picture of interoperability.

Reference Model

Initially *STACKS* begins with the OSI model and Malamud's revision, in which three new levels are layered on top: finance, politics, and religion. (The upper two layers are probably indistinguishable to the untrained eye, but this is a matter for future historians to decide.) Following this, we launch into a discussion as to why real-world solutions are largely multi-protocol. To present this, Malamud uses his previously published treatise on "Mangoes and Orangutans" in which he argues that of the three big-ticket network file methods, FTP, NFS, nor FTAM, none is ideal in all circumstances. In brief, each has evolved with a particular service model in mind, and, when used in combination, they can be complimentary in nature. [Ed.: See *ConneXions,* Volume 4, No. 11, November 1990 and *ConneXions,* Volume 5, No. 4, April 1991.]

New technologies

Following this, the concept of networks and internets in the real world is developed. Malamud makes the point, and rightly so, that its difficult to figure out the boundaries of our networks. The days of "our network is this yellow cable in this room" have disappeared. This leads to a discussion of the new technologies being developed for:

- Local area networks, i.e., FDDI, HIPPI—a High Performance Parallel Interface operating at 800 or 1600Mbps; and,
- Wide area networks, i.e., the ever-elusive ISDN, the amazingly popular Frame Relay, and some other technologies such as ATM, SONET, and my personal favorite, SMDS

Abstractions

STACKS then moves towards the edges of the network as it looks at how protocol suites are put together on host systems. The discussion focuses on two different abstractions used to hide the transport layer from the application: AT&T's *STREAMS* and DEC's *Towers.* In both cases, these are architectural mechanisms used to organize software—the absence or presence of such an abstraction shouldn't change the bits exchanged by the protocols.

The author is a bit light in this area as he doesn't really get into the weaknesses of these kind of schemes (in contrast, in earlier sections he is conversant where other technologies start to get into trouble). Once the discussion of *STREAMS* and *Towers* is out of the way, he looks at some interesting work being done in the end-to-end area:

- Speed (how fast can data be reliably moved between hosts); and,
- Size (how many hosts can be attached to an internet).

Glue

These scaling issues are then followed with a look at the "glue" which is used to bind networks together: routing protocols. Once the painful topic of so-called "extended Ethernets" (using bridges rather than routers to join remote LANs), STACKS looks at dynamic routing mechanisms. After a brief discussion on the theory of routing protocols, the current generation of protocols are briefly introduced: OSI's IS–IS, OSPF (the Internet's *Open Shortest Path First*), the old EGP (*Exterior Gateway Protocol*), and finally BGP (the *Border Gateway Protocol*).

Open Wars

STACKS then takes a look at the "Open Wars" as exemplified by the myriad of organizations competing to define open systems: OSF, ONC, OSI, etc. (Reviewer's comment: guess what word the first letter of each acronym stands for—now the fun part, observe that the word has entirely different meaning based on the acronym in which it appears). The power of *STACKS* in examining these efforts is showing how they contain many of the same components (e.g., ONC and OSF use MIT's *X Window System,* and OSI is considering it), and yet are miles apart on other issues such as naming. Finally, this section wraps up by looking at how some vendors are trying to reconcile these issues in their individual product lines.

The cost of standards

STACKS then wraps up with a discussion on network security (largely devoted to public key cryptosystems and the first large-scale use of such a technology in the Internet—*Privacy Enhanced Mail*), followed by a brief introduction to some efforts looking into gigabit networking and digital libraries. But, the best is saved for last as Malamud goes after the Public Standards Cartel with a vengeance. This closing chapter in *STACKS* is simply a classic: in 20 pages or so Malamud explains why organizations like the ISO and CCITT have failed to produce useful public standards—it is simply too difficult and expensive to get copies of the damnable things. There are too many good lines to repeat, but here is one of my favorites: Malamud orders up copies of some standards from a US supplier, the total cost is US$1350 and the cost per inch is—can you believe it—US$388! Malamud's point, made deftly at the end of the chapter is that the usefulness of a standard increases when anyone can get a copy, study it, and then implement it.

But, perhaps my favorite of Malamud's anecdotes is what happened when he sent copies of a draft of *STACKS* to several people, asking for a possible technical review (a commonly accepted practice for writers of professional texts). Speaking from experience, when you get back responses, the feedback is usually quite good. Well, one copy was sent to the OSF. This was completely reasonable as *STACKS* contains a fair bit of material on the OSF and technologies being developed by the OSF. Now, imagine Malamud's surprise when he gets back his manuscript from the OSF, along with a letter from their *General Counsel* indicating that they have returned the draft untouched because they don't review materials submitted by outside parties, in the interest of "vendor neutrality."

continued on next page

Book Review *(continued)*

This legal type then warns Malamud not to represent that the OSF endorses his book in any way. Here Malamud is giving the OSF a chance to clear up any misunderstandings he might have about their work, and they drop a lawyer, a General Counsel no less, on him. I guess professional courtesy goes by the wayside when you're dealing with the big business of open systems. Of course, given the composition of the OSF, its membership and licensing arrangements, it can hardly claim to be "vendor neutral."

The Salman Rushdie Effect

Being good-natured, Malamud takes this in a humorous vein postulating it might have a Salman Rushdie effect: "Buy this book, OSF didn't endorse it." Well, I'd buy the book anyway—it's a great book (it's also got an excellent 40 page glossary that I forgot to mention earlier). But now, I'm going to buy two copies of *STACKS*. My second copy I'm going to send to the OSF, with a cover letter telling them what a great book it is! I suggest you do the same.

—Marshall T. Rose

Network Reading List available

by Charles Spurgeon, University of Texas at Austin

A new version of the document "Network Reading List: TCP/IP, UNIX, and Ethernet" is now available from the Network Information Center at the University of Texas at Austin. The list may be found on host `ftp.utexas.edu` (128.83.185.16). This is version 3.0 of the reading list, dated August, 1991.

What is it?

The network reading list is an annotated list of books and other resources focusing on three networking technologies that are in wide use: TCP/IP, UNIX, and Ethernet. A mix of resources is presented ranging from introductory information to in-depth technical details. Version 3.0 of the list has been completely rewritten and updated, and now includes nearly 70 items. The list is weighted towards resources that cover the territory well, and that deal with real-world problems found on growing networks. The table of contents is included below.

How do I get a copy?

You can retrieve a copy of the list in either *PostScript* format or as a plain ASCII text file. The *PostScript* format is recommended. The *PostScript* file is 34 pages long, and the ASCII text file contains 51 pages. Copies of the list may be retrieved using anonymous FTP or a mail-based archive server program.

The hostname for anonymous FTP is `ftp.utexas.edu`, and the files are in the `pub/netinfo/docs` and `pub/netinfo/ps` directories as `net-read.txt` and `net-read.ps` respectively.

The e-mail address of the archive server program is `archive-server@ftp.utexas.edu`. You can retrieve copies of the list by sending the archive server program a command line in the body of an electronic mail message. The command line:

```
send ps net-read.ps
```

will cause the archive server program to send you a copy of the *PostScript* file, while the command line:

```
send docs net-read.txt
```

will retrieve the ASCII text file.

The command line should be placed in the body of a message sent to `archive-server@ftp.utexas.edu`. The archive server is just a simple program, so make sure to send the command line exactly as shown here.

Table of contents

Section 1 — TCP/IP:

Section 2 — UNIX:

Section 3 — Ethernet:

Section 4 — Interest Groups, Periodicals, and Conferences:

Charles Spurgeon
Computation Center
Networking Services
University of Texas at Austin
Room COM1
Austin TX 78712
c.spurgeon@utexas.edu
(512) 471-3241 ext 265

CONNEXIONS
480 San Antonio Road
Suite 100
Mountain View, CA 94040
415-941-3399
FAX: 415-949-1779

ADDRESS CORRECTION
REQUESTED

FIRST CLASS MAIL
U.S. POSTAGE
PAID
SAN JOSE, CA
PERMIT NO. 1

CONNEXIONS

EDITOR and PUBLISHER — Ole J. Jacobsen